The role of edge-based and surface-based information in natural scene categorization: evidence from behavior and event-related potentials

Article (Accepted Version)

The role of edge-based and surface-based information in natural scene categorization: Evidence from behavior and event-related potentials

Qiufang Fu[1], Yong-Jin Liu[2], Zoltan Dienes[3], Jianhui Wu[1], Wenfeng Chen[1], Xiaolan Fu[1],

[1] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

[2] Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

[3] Sackler Centre for Consciousness Science and School of Psychology, University of Sussex, BN1 9QH, Brighton, United Kingdom

Address for correspondence:

Qiufang Fu Ph. D

Institute of Psychology, Chinese Academy of Sciences

16 Lincui Road, Chaoyang District, Beijing 100101, China

Tel: (86 10) 6484-5395

Fax: (86 10) 6487-2070

E-mail: fuqf@psych.ac.cn

Abstract

A fundamental question in vision research is whether visual recognition is determined by edge-based information (e.g., edge, line, and conjunction) or surface-based information (e.g., color, brightness, and texture). To investigate this question, we manipulated the stimulus onset asynchrony (SOA) between the scene and the mask in a backward masking task of natural scene categorization. The behavioral results showed that correct classification was higher for line-drawings than for color photographs when the SOA was 13 ms, but lower when the SOA was longer. The ERP results revealed that most latencies of early components were shorter for the line-drawings than for the color photographs, and the latencies gradually increased with the SOA for the color photographs but not for the line-drawings. The results provide new evidence that edge-based information is the primary determinant of natural scene categorization, receiving priority processing; by contrast, surface information takes longer to facilitate natural scene categorization.

Key words: natural scene categorization, edge-based theory, surface-based theory, SOA, ERPs

1. Introduction

Humans have a remarkable ability to categorize natural scenes quickly and accurately. The human brain needs only approximately 150 ms to decide whether a color photograph, flashed for 20 ms, contains animals or vehicles (Rousselet~~ et al.~~, Fabre-Thorpe, & Thorpe, 2002; Thorpe~~ et al.~~, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001), even with little or no attention applied to the task (Feifei, VanRullen, Koch, & Perona~~ et al.~~, 2005; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002; Li, VanRullen, Koch, & Perona~~et al.~~, 2002). The challenge is to explain how rapid natural scene categorization takes place in the human brain.

A recent fMRI study found that line-drawings generated similar neural activation as color photographs in the parahippocampal place area (PPA) and the retrosplenial cortex (RSC), which suggests that the human visual system uses schematic representations with content that is analogous to simple line-drawings, to encode and process statistical regularities in a scene (Walther~~ et al.~~, Chai, Caddigan, Beck, & Fei-Fei, 2011). This finding has provided new evidence for an edge-based theory that assumes that edge-based representations are sufficient for object recognition and that surface characteristics such as color, brightness, and texture are less efficient routes for accessing the memorial representation (Biederman, 1987; Bie~~i~~derman & Ju, 1988). Indeed, some studies have found that surface gradients such as color changes had little influence on object classification and identification (e.g., Bie~~i~~derman & Ju, 1988; Cave, Bost, & Cobb~~et al.,~~, 1996; Joseph & Proffitt, 1996) or even impaired object classification (e.g., Gagnier & Intraub, 2012). For example, Bie~~i~~derman and Ju (1998) demonstrated that the reaction times and error rates were virtually identical for the common objects of color photographs and line-drawings when the images were briefly (50-100 ms) presented. Thus, although scene recognition and object recognition are technically different, the same perceptual processes might be involved. However, due to poor temporal resolution, on the order of one to several seconds (Rossion, Kung, & Tarr~~n et al.~~, 2004), the above fMRI study cannot discriminate differences in the time course of categorizing color photographs and line-drawings.

The role of surface properties in object or scene recognition remains controversial (e.g., Gagnier & Intraub, 2012; Parron & Washburn, 2010; Wichmann, Sharpe, & Genenfurtner~~ et al.~~, 2002). In contrast with the edge-based theory, the alternative surface-based theory assumes that surface gradients are central for object recognition and that both contour and surface information provide simultaneous routes for basic-level categorization. This perspective has received support from other studies (e.g., Tanaka~~ et al.~~, Weiskopf, & Williams, 2001; Wichmann et al., 2002; Wurm, Legge, Isenberg, & Luebker,~~et al.,~~ 1993). For example, color improved object recognition of common food items when there was no time limit on the stimulus presentation (Wurm et al., 1993).

Interestingly, Laws and Hunter (2006) did not find a significant difference in the accuracy between the objects in color photographs and line-drawings with a 20-ms presentation of each image, which is consistent with the findings of Bie~~i~~derman and Ju (1988), but a marginally significant advantage for color photographs over line-drawings was found (p = .07) with a 1000-ms presentation of each image, which is principally consistent with the findings of Wurm et al. (1993). A comparison of the above studies also reveals that most of the studies in support of the edge-based theory limited the presentation times or processing duration to a very short time, while there was no time limit or a long processing time in the study that supported the surface-based theory. Thus, we predict that the stimulus presentation or processing duration could modulate the role of the surface information in scene perception. Specifically, if the processing duration is long enough, then the surface information should facilitate the recognition; however, if the processing duration is extremely short, then surface information could even impair recognition performance if edge-based information is thereby harder to extract. If the latter occurs, then the result will provide new evidence for edge-based information receiving priority processing.

The purpose of the present study was to address this issue by adopting event-related potentials (ERPs) in a backward masking task of categorizing natural scenes. To manipulate the processing duration, a backward masking paradigm was adopted in the present ERP study, in which the stimulus duration was constant but the stimulus onset asynchrony (SOA) between the image and mask was varied. Backward masking is useful in investigating the time course of information processing in the visual system in that it allows processing to be interrupted at different times (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe~~ et al.~~, 2005; Hansen & Loschky, 2013; Holcomb & Grainger, 2006; Kovács, Vogels, & Orban~~acs et al.~~, 1995; Loschky et al., 2007; Loschky, Hansen, Sethi, & Pydimarri~~et al.~~, 2010; Macknik & Livingstone, 1998; Rieger, Braun, Bülthoff, & Gegenfurtner,~~ et al.,~~ 2005; Rolls~~-~~, Tovée, & Panzeri~~et al.~~, 1999; VanRullen & Koch, 2003). Usually, when the SOA becomes longer, the behavioral performance and neural activation recorded by fMRI increase and so does the ERP differential activity, roughly between 150 and

250 ms on the targets and distracters (Bacon-Macé et al., 2005; Holcomb & Grainger, 2006). Because accuracy increases significantly with SOAs below 44 ms (i.e., 6.25, 12.50, 18.75, 25, 31.25, 43.75 ms, see Bacon-Macé et al., 2005), the SOA was set at 13, 27, 40 and 213 ms in the present study. Moreover, to explore the role of edge-based and surface-based information in natural scene categorization, we adopted color photographs and line-drawings as stimuli because the color photographs include both edge-based information (e.g., edge, line, and conjunction) and surface properties (e.g., color, brightness, and texture), whereas line-drawings include only edge-based information, as established in a previous study (Walther et al., 2011).

Previous ERP studies have shown that natural scene categorization involves two stages: a perception stage that extracts information about different features of the visual input and a decision stage that evaluates the relevance of the information in making a decision (VanRullen & Thorpe, 2001; Bacon-Macé et al., 2005). Early ERP components such as P1 and N1 are associated with feature detection or integration (Hillyard & Münte, 1984) and are sensitive to elemental features of stimuli (e.g., Holcomb~~er~~ & Grainger, 2006; Itier, Latinus, & Taylor ~~et al.~~, 2006). Given that color photographs involve both edge-based and surface-based information while line-drawings include only edge-based information, differences in the components at the perceptual stage could be elicited by color photographs versus line-drawings. Specifically, if surface-based information is processed simultaneously with edge-based information such that both facilitate categorization, then the latencies of the early components should be at least as fast for the color photographs as for the line-drawings. Conversely, if the latencies of early components are faster for the line-drawings than the color photographs, then it indicates the importance of edge-based information, with surface information analyzed as a secondary route for visual cognition which does not facilitate early on. Indeed, Walther et al. (2011) found that there was only a low correlation between the neural activity that was generated by color photographs and the neural activity that was generated by line-drawings in early visual areas, which suggests that the feature analysis in early visual processing differs between color photographs and line-drawings. The later ERP components, such as N2 and P3, are related to decision making (Folstein & Van Petten, 2008; Nieuwenhuis~~ et al.~~,~~–~~ Aston-Jones, & Cohen, 2005). If only an edge-based representation is sufficient in the decision-making process, then there will be no difference in the pattern of later components of color photographs and line-drawings, as suggested by Walther et al. Conversely, if an edge-based representation is not sufficient in the decision-making process, then there will be differences in the pattern of later components between color photographs and line-drawings.

To the best of our knowledge, although a considerable number of studies have investigated the role of surface information in scene and object recognition (Bie~~i~~derman & Ju, 1988; Delorme~~-~~, Richard, & Fabre-Thorpe~~et al.~~, 2000; Gagneier & Intraub, 2012; Walther et al., 2011), few studies have adopted ERP techniques to address this question. Goffaux et al. (2005), using a go/no-go paradigm, measured ERPs when people categorized normally colored, grayscale and abnormally colored scenes with a 100-ms presentation. They found that the reaction times and accuracy were optimal for the normal version, followed by the grayscale and then the abnormal version; the onset of the early ERP component at the frontal sites mirrored these effects. Thus, color contributes to rapid natural scene categorization, which is consistent with the surface-based account. However, although the stimulus presentation in this study was 100 ms, they used a long variable SOA of 1500 to 1800 ms. Because no masks were used to interrupt the processing, there was sufficient time to process the image fully. Moreover, their concern was the role of color in natural scene categorization, and thus, they did not answer how edge-based and surface-based information contributes to natural scene categorization. The role of edge-based and surface-based information and its interaction with the processing duration in scene recognition remains an open question.

In addition, all of the above ERP studies used a go/no-go paradigm, during which people first made a decision about whether the image contained animals or vehicles and then performed the go or no-go reaction. Because the targets and distracters belonged to different categories, the differential activity between the targets and distracters might have reflected a difference in either a high-level property such as the category or a low-level property such as the contrast (Rousselet~~t~~ & Pernet, 2011). To avoid this ambiguity, we used a forced-choice rather than go/no-go task and compared the differential activity between incorrect and correct trials. Because there were incorrect and correct trials for each category, the analysis of the differential activity according to the correctness should reflect how people correctly categorize scenes. Second, because the targets and distracters require different responses in the go/no-go task, the differential activity between the targets and distracters could result from either different decision-making processes or different preparations for reactions. To dissociate the reaction preparations from the decision-making processes, the locations of six category names were randomly assigned on each trial, and a blank was displayed for 500 ms before the presentation of the category names (see Figure 1).

2. Experimental procedures

2.1 Participants

Twenty-two undergraduate and graduate students (11 male, 11 female), aged 19-29 years (*M* = 21.82, *SD* = 2.34), voluntarily took part in this experiment and were paid for their attendance. All of them had normal or corrected-to-normal vision and gave the written informed consent. None of them had any history of neurological or psychiatric diseases. This experiment was conducted in accordance with the Declaration of Helsinki and was approved by the committee for the protection of subjects at the Institute of Psychology, Chinese Academy of Sciences.

2.2 Materials

Color photographs and line-drawings of six natural scene categories (beaches, city streets, forests, highways, mountains and offices) were adopted as stimuli, which were first used by Walther et al. (2011)[1]. Each category had 76 to 80 different images, for a total of 475. Each image had two versions: one was a color photograph, and the other was a line-drawing. The line-drawings were produced by trained artists by tracing the contours in the color photographs (see Walther et al., 2011). All of the images were resized to 320 * 240 pixels. Two white noise images at two different spatial scales were generated as masks: one was generated at the resolutions of 320*240, and the other was generated at the resolutions of 20*15 and then resized to 320 * 240 pixels. Each mask also had two versions: one in color, the other in grayscale. The experiment was carried out on a CRT monitor with a resolution of 1280 * 768 pixels, a mean luminance of 50.6

[1]. We got all of the images from Dr. Walther.

cd/m^2, and a refresh rate of 75 Hz. The images of six natural scene categories were presented on a silver gray background with a mean luminance of 27.4 cd/m^2.

2.3 Procedure

The participants were seated in an electrically shielded, dimly lit and sound-attenuated room. ~~The images of six natural scene categories were presented on a monitor with a resolution of 1280 * 768 pixels and a refresh rate of 75 Hz.~~ The distance from the participants' eyes to the center of the screen was approximately 60 cm when they sat straight in the chair, and no chinrest was used. Each image was approximately 8.70 degree wide and 8.19 degree high. The participants were tested for 10 blocks, for a total of 950 trials. At the beginning of each trial, a black fixation cross was presented on a ~~neutral~~ ~~grey~~ silver gray background in the center of the screen for 500-950 ms at random (see Figure 1). Then, an image was flashed for 13 ms, which was followed by two masks[2]. Each mask was shown for 50 ms, for a total of 100 ms. The sequence of the two masks was randomly assigned on each trial. After the masks, there was a blank of 500 ms. The stimulus onset asynchrony (SOA) between the image and the mask was 13, 27, 40 or 213 ms, at random. After the blank, six category names appeared on the screen from left to right, for which the locations corresponded to the keys D, F, G, H, J, and K on the keyboard. On each trial, the locations of the six category names were randomly assigned, and thus, the participants would not prepare their response before its appearance and no response bias toward a favored location

[2] In a preliminary experiment, we found that the mask with a high resolution masked color photographs (non-significantly) more than line-drawings, while a mask with a low resolution masked line-drawings (non-significantly) more than color photographs. To balance possibly different masking effects on color photographs and line-drawings, we used the two masks.

would contaminate the results[3] (Loschky, Ringer, Ellis, & Hansen, 2015). The participants were forced to make a choice among the six categories by pressing the corresponding key, and there was no time limit for them to make the choice. There was no feedback about the correctness of their response. After their response, they were further asked to report "how clearly did you see the image" with four possible responses from left to right on a perceptual awareness scale (PAS), by pressing the corresponding key. Then, the participants were asked to press the space key to begin the next trial when they were ready. In each block, half of the images were color photographs, and half were line-drawings, with equal trials in each category and each SOA. There was at least a 30-second rest between any two blocks.

2.4 EEG recording and analysis

The EEG was recorded from 64 scalp sites using Ag-AgCl electrodes in an elastic cap according to the International 10-20 system. The vertical and horizontal EOG were recorded with two pairs of electrodes placed 1 cm above and below one eye and 1 cm lateral from the outer canthi of both eyes. The left mastoid was used as an on-line reference, and the algebraic average of the left and right mastoids was used as an off-line re-reference. The EEG and EOG signals were amplified by a NeuroScan Synamps amplifier with a band pass of 0.05–100 Hz and digitized at 500 Hz. The impedance of the electrodes was maintained below 5 kΩ. EEG data were low-pass filtered with a cutoff frequency at 30 Hz and averaged offline for epochs of 800 ms, starting 100 ms prior to the stimulus onset and ending 700 ms afterward. A baseline correction

[3]. However, as participants need first search the target category name on each trial, thus the reaction time was inflated and the accuracy rather than RT was used as a dependent variable.

was performed for each epoch using the 100 ms before the presentation of each image. Trials with artifacts that were determined by a criterion of 80 μV were rejected offline, which amounted to only 2.6% of the trials.

The ERPs were first averaged separately across correct and incorrect trials for each type of image and SOA for each subject. The SOA of 213 ms was not included because of having an insufficient number of incorrect trials for the ERP average. In the statistical analysis of the ERP data, we focused on the peak latencies and amplitudes of the posterior P1 (80-160 ms) and N1 (130–210 ms) and anterior P2 (140-240 ms) and the mean amplitudes of the posterior P2 (210–260 ms) and anterior N2 (220–320 ms), P3 (370–420 ms), and N4 (420–520 ms). The time windows were chosen because they best captured the differences among the different conditions and were relatively free from overlap with adjacent ERPs. Based on previous studies (e.g., Bacon-Mace et al., 2005; Melloni et al., 2011) and the topography of each component in the present study, a group of occipital electrodes (CB1, O1, Oz, O2, and CB2) was selected for the posterior P1, N1, and P2 (see Figures 3C, 3E); a group of fronto-central electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4) was selected for the anterior P2, N2, P3, and N4 (see Figures 5C, 5E). The latencies and amplitudes were computed as the means over groups of electrodes that were representative of the topography of each component for each subject. The mean and standard error for each component were computed across subjects. Each was subjected to a repeated measures three-way ANOVA with the factors of type of image (color photographs vs. line-drawings), SOA (13 ms vs. 27 ms vs. 40 ms), and correctness (correct vs. incorrect).

Key non-significant results were interpreted with Bayes factors. P-values by themselves cannot discriminate insensitive data from support for the null hypothesis, whereas Bayes factors make that distinction. More specifically, when using the Bayes factor, B, to compare an alternative hypothesis (H1) against the null hypothesis (H0), if B is greater than 3, then there is substantial evidence for H1 over H0; if B is less than 1/3, then there is substantial evidence for H0 over H1; and if B is between 3 and 1/3, then the data do not discriminate H0 from H1 (Dienes, 2011). Bayes factors were determined using the free online software associated with Dienes (2014), which is located at http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayes_factor.swf, with the Matlab and R code provided at http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm. Dienes (2014) provides a tutorial.

3. Results

3.1 Behavioral results

Figure 2A shows the accuracy rate for color photographs and line-drawings for each SOA. Because the task was to make a choice among the six categories, the chance probability is 0.17. The accuracy for the color photograph trials was significantly better than chance for each SOA [SOA = 13 ms: t (21) = 11.08, $p < .001$, dz = 2.36; SOA = 27 ms: t (21) = 14.51, $p < .001$, dz = 3.09; SOA = 40 ms: t (21) = 20.52, $p < .001$, dz = 4.38; SOA = 213 ms: t (21) = 75.99, $p < .001$, dz = 16.20], as was the accuracy for the line-drawing trials for each SOA [SOA = 13 ms: t (21) = 10.66, $p < .001$, dz = 2.27; SOA = 27 ms: t (21) = 12.96, $p < .001$, dz = 2.76; SOA = 40 ms: t (21) = 16.31, $p < .001$, dz = 3.47; SOA = 213 ms: t (21) = 51.09, $p < .001$, dz = 10.89]. The results suggested that people could correctly classify natural scenes for each SOA. A repeated ANOVA with the type of image and SOA as within-subject factors revealed that overall, the accuracy rate was higher for the color photograph trials than for the line-drawing trials [.66 ± .02 vs. .62 ± .02, F (1, 21) = 18.87, p < .001, η_p^2 = .47], which increased with SOA [F (1.95, 41.00) = 261.20, p < .001, η_p^2 = .93, using the Greenhouse-Geiser correction], and the increase with the SOA was influenced by the type of image [F (3, 63) = 14.42, p < .001, η_p^2 = .41]. Further analysis revealed that the accuracy rate was significantly higher for the color photograph trials than for the line-drawing trials when the SOAs were 27, 40 and 213 ms [SOA = 27 ms: .61 ± .03 vs. .55 ± .03, t (21) = 3.62, p < .01, dz = .77; SOA = 40 ms: .72 ± .03 vs. .64 ± .03, t (21) = 4.81, p < .001, dz = .96; SOA = 213 ms: .90 ± .01 vs. .84 ± .01, t (21) = 5.57, p < .001, dz = 1.19], but was lower for the color photograph trials than for the line-drawing trials when the SOA was 13 ms [.41 ± .02 vs. .44 ± .03, t (21) = -2.16, p < .05, dz = .46]. The results indicated that the facilitatory role of surface information in natural scene categorization is modulated by the processing duration. A lower performance for the color photograph trials than for the line-drawing trials when SOA was 13 ms revealed that surface-based information could impair recognition performance when the processing time was extremely limited, providing evidence for edge-based information receiving priority processing.

To further examine the contribution of surface-based and edge-based information to accuracy, we took the accuracy difference between the color photograph trials and the line-drawing trials as the accuracy contributed by surface properties and took the accuracy difference between the line-drawing trials and chance level (i.e., .17) as the accuracy contributed by edge-based features (see Figure 2B). A repeated ANOVA with the contribution of different types of information and SOA as within-subject factors revealed that overall the accuracy contributed by edge-based information was much larger than that contributed by surface-based information [.04 ± .01 vs. .45 ± .02, F (1, 21) = 191.36, p < .001, $\eta_p{}^2$ = .90], the accuracy contributed by edge-based and surface-based information increased with SOA [F (1, 21) = 212.77, p < .001, $\eta_p{}^2$ = .91], and the increase with SOA was influenced by the contribution type [F (1, 21) = 34.93, p < .001, $\eta_p{}^2$ = .63]. Further analysis revealed that the accuracy contributed by the edge-based information was much larger than that contributed by the surface-based information for each SOA [SOA = 13 ms: -.04 ± .02 vs. .27 ± .03, t (21) = 8.01, p < .001, dz = 1.71; SOA = 27 ms: .06 ± .02 vs. .38 ± .03, t (21) = 8.24, p < .001, dz = 1.76; SOA = 40 ms: .08 ± .02 vs. .47 ± .03, t (21) = 9.81, p < .001, dz = 2.09; SOA = 213 ms: .07 ± .01 vs. .67 ± .01, t (21) = 25.39, p < .001, dz = 5.41]. Interestingly, the contribution of edge-based information gradually and significantly increased with SOA (all ps < .001), whereas the contribution of surface-based information increased from SOA of 13 ms to SOA of 27 ms [-.04 ± .02 vs. .06 ± .02, t (21) = 5.17, p < .001, dz = 1.10], but there were no significant difference among SOAs of 27, 40, and 213 ms (all ps > .34). The results indicated that the edge-based information plays a primary role and the surface-based information a secondary role in natural scene categorization.

Finally, we calculated the average awareness score for each SOA of color photographs and line-drawings (see Figure 2C). When SOA was 13 ms, the awareness scores were significantly above 1 (no experience) for both types of images [color photographs: t (21) = 8.97, p < .001, dz = 1.91; line-drawings: t (21) = 9.14, p < .001, dz = 1.95], but were not significantly different from 2 (brief glimpse) [color photographs: t (21) = -.61, p = .55; line-drawings: t (21) = -.21, p = .83]. When SOA was 27 ms, the awareness score for color photographs was significantly above 2 (weak glimpse) [t (21) = 3.38, p < .01, dz = .72], but significantly below 3 (almost clear experience) [t (21) = -5.45, p < .001, dz = 1.12]; the awareness score for line-drawings was not significantly above 2 (weak glimpse) [t (21) = 1.70, p = .10], and significantly below 3 (almost clear experience) [t (21) = -7.32, p < .001]. When the SOA was 40 ms, the awareness score for both types of images were significantly above 2 (weak glimpse) [color photographs: t (21) = 5.68, p < .001, dz = 1.21; line-drawings: t (21) = 3.71, p = .001, dz = .79], but significantly below 3 (almost clear experience) [color photographs: t (21) = -2.62, p < .05, dz = .56; line-drawings: t (21) = -4.88, p < .001, dz = 1.04]. When the SOA was 213 ms, the awareness score for color photographs was significantly above 3 (almost clear experience) [t (21) = 4.29, p < .001, dz = .91], but significantly below 4 (clear experience) [t (21) = -6.48, p < .001, dz = 1.38]; the awareness score for line-drawings was not significantly above 3 (almost clear experience) [t (21) = .85, p = .40], and significantly below 4 (clear experience) [t (21) = -9.23, p < .001, dz = 1.97]. That is, participants reported having experience below “almost clear experience” for both types of images when SOA were 13, 27, and 40 ms, and having mainly “almost clear experience” only when SOA was 213 ms.

3.2 ERP results

The ERP data of the color photographs and line-drawings in both the correct and incorrect trials at the occipital sites (CB1, O1, Oz, O2, and CB2) and fronto-central sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4) were analyzed when the SOA was 13, 27, and 40 ms. The SOA of

213 ms was not included because of having an insufficient number of incorrect trials for the ERP average. We first consider how the type of image, SOA, and correctness influenced the posterior P1, N1, and P2 at the occipital sites. Then, we show how the factors modulated the anterior P2, N2, P3, and N4 at the fronto-central sites. Three-way repeated ANOVAs with the type of image, SOA, and correctness as within-subject factors were performed over the latencies or amplitudes of each component. To demonstrate the different time courses of the natural scene categorization of the color photographs and line-drawings, we reported only two-way interactions between the type of image and the SOA and between the type of image and the correctness. Finally, we will explore the relationship between the behavioral accuracy and ERP effects by using regression analysis.

3.2.1 Posterior P1, N1, and P2 effects

Figure 3 shows ERP data at the occipital electrode sites, in which the ERP waveforms were computed over the group of occipital electrodes (CB1, O1, Oz, O2, and CB2), which was representative of the topography of each component. Figure 4 shows the latencies or amplitudes of the posterior P1, N1, and P2 under each condition. Table 1 summarizes the significant results of the three-way repeated ANOVAs that were performed over the latencies or amplitudes of the posterior P1, N1, and P2.

Peak latencies of the posterior P1 and N1. For the P1 peak latencies, there was only a significant SOA by the type of image interaction. As shown in Figure 4, consistent with the edge-based theory, the P1peak latency was significantly shorter for the line-drawing trials than for the color photograph trials when SOA was 40 ms [122.61 ± 3.43 ms vs. 115.19 ± 3.30 ms, t (21) =

2.39, $p < .05$, $dz = .51$]; but inconsistent with the edge-based theory, the P1 peak latency was significantly shorter for the color photograph trials than for the line-drawing trials when SOA was 13 ms [110.61 ± 2.19 ms vs. 116.29 ± 2.67 ms, $t(21) = -2.22$, $p < .05$, $dz = .47$], and there were no significant differences between the color photograph trials and the line-drawing trials when SOA was 27 ms [118.43 ± 2.57 ms vs. 118.45 ± 3.07 ms, $t(21) = -.01$, $p = .99$]. However, more importantly, the P1 peak latency significantly increased with the SOA only for the color photograph trials ($ps < .05$) but not for the line-drawing trials ($ps > .12$). To interpret the latter non-significant results, Bayes factors were used (Dienes, 2011). Nothing at all follows from a non-significant result in itself, but a Bayes factor (B) can indicate substantial evidence for the null hypothesis ($B < 1/3$), that the data are insensitive ($1/3 < B < 3$), or substantial evidence for the alternative ($B > 3$). Because the linear trend was significantly greater for the color photograph trials than for the line-drawing trials, the alternative hypothesis for the line-drawing trials can be represented as being uniform between 0 and the maximum provided by the linear trend estimated for the color photograph trials. For the P1 latencies, the linear trend for the line-drawing trials was -1 ms (SE = 2 ms); using the uniform range [0, 12] to represent the alternative (where 12 was the linear trend for color photographs) yields $B = 0.15$. In other words, there is substantial evidence for the null hypothesis of no linear trend in the P1 latencies for the line-drawing trials over the alternative. Thus, the results indicated that shorter SOA was sufficient for extracting information from line-drawings rather than color photographs, which was principally consistent with the hypothesis of the edge-based theory.

For the N1 peak latencies, there was a significant type of image by SOA interaction. As shown in Figure 4, consistent with the edge-based theory, there was significantly shorter N1 peak latency for the line-drawing trials than for the color photograph trials when SOA was 27 and 40 ms [SOA = 27 ms: 172.39 ± 1.87 ms vs. 162.20 ± 2.20 ms, t (21) = 5.37, $p < .001$, $dz = 1.15$; SOA = 40 ms: 180.49 ± 1.85 ms vs. 161.21 ± 3.46 ms, t (21) = 5.70, $p < .001$, $dz = 1.21$], but there were no significant differences on the N1 peak latency between color photograph trials and line-drawing trials when SOA was 13 ms [162.93 ± 1.78 ms vs. 159.74 ± 2.54 ms, t (21) = 1.55, $p = .14$]. However, more importantly, the N1 peak latency significantly increased with SOA for the color photograph trials ($ps < .05$) but not for the line-drawing trials ($ps > .17$). Similarly, the linear trend was 1 ms ($SE = 3.5$ ms) for the line-drawing trials; using the uniform [0, 18] to represent the alternative (where 18 was the linear trend for the color photographs) yields $B = 0.31$, which is also substantial evidence for the null hypothesis. The results confirmed that shorter SOA was sufficient for extracting information from line-drawings rather than color photographs, which was substantially consistent with the edge-based theory.

In addition, the type of image by correctness interaction was also significant. The N1 peak latency was significantly shorter for the incorrect than correct trials for the color photograph trials [168.87 ± 1.64 ms vs. 175.01 ± 2.03 ms, t (21) = -4.10, $p = .001$, $dz = .87$], but not for the line-drawing trials [162.19 ± 2.41 ms vs. 159.91 ± 2.59 ms, t (21) = 1.77, $p = .09$]. That is, the N1 peak latency was related to correct classification for color photographs.

Amplitudes of the posterior P1, N1, and P2. For the posterior P1 peak amplitudes, there was a significant type of image by SOA interaction. For both types of images, the P1 amplitude significantly increased from a SOA of 13 to a SOA of 27 [color photographs: 2.77 ± .39 μV vs.

3.58 ± .45 μV, t (21) = -2.90, $p < .01$, dz = .62; line-drawings: 1.65 ± .40 μV vs. 2.28 ± .51 μV, t (21) = -2.56, $p < .05$, dz = .55], but not from a SOA of 27 to a SOA of 40 (ps > .29). The interaction of the type of image by correctness also reached significance. For the color photographs, the P1 amplitude was significantly larger for correct than incorrect trials for the color photographs [3.54 ± .41 μV vs. 3.22 ± .44 μV, t (21) = 3.51, $p < .01$, dz = .75], but the P1 amplitude was marginally significantly smaller for the correct than the incorrect trials for the line-drawings [1.73 ± .49 μV vs. 2.08 ± .44 μV, t (21) = -2.04, p = .054, dz = .43]. That is, incorrect classification was related to different P1 effects for color photographs and line-drawings.

For the posterior N1 peak amplitudes, there was a significant type of image by correctness interaction. The N1 effect was significantly larger for the incorrect than correct trials for the color photographs [-5.71 ± .48 μV vs. -4.64 ± .52 μV, t (21) = 4.51, $p < .001$, dz = .96] but not for the line-drawings [-3.03 ± .45 μV vs. -3.02 ± .46 μV, t (21) = .08, p = .94]. Furthermore, for color photographs, the comparison between incorrect and correct trials for each SOA revealed that the N1 effect was significantly larger for the incorrect than correct trials when the SOA was 27 and 40 ms [SOA = 27 ms: -5.86 ± .59 μV vs. -4.54 ± .53 μV, t (21) = 4.35, $p < .001$, dz = .93; SOA = 40 ms:-5.89 ± .57 μV vs. -4.03 ± .58 μV, t (21) = 4.54, $p < .001$, dz = .97, respectively], but not when the SOA was 13 ms [-5.37 ± .42 μV vs. -5.34 ± .53 μV, t (21) = .08, p = .94]. The difference for SOA of 13 ms was -.03μV (SE = .37), using the uniform [-1.86, 0] to represent the alternative (where -1.86 was the difference for SOA of 40) yields B = 0.27, which is substantial evidence for the null hypothesis. Thus, the results suggested that N1 was related to correct classification of color photographs when SOA was longer than 13 ms.

For the posterior P2 amplitudes, there was a significant type of image by SOA interaction. The posterior P2 amplitude was significantly larger for the line-drawing trials than for the color photograph trials for each SOA (*p*s < .001), while the P2 amplitude significantly decreased with the SOA for both types of images (*p*s < .001). The interaction of the type of image and correctness was also significant. For color photographs, the P2 amplitude was significantly larger for incorrect than correct trials [2.68 ± .41 μV vs. 1.82 ± .36 μV, t (21) = 3.96, p = .001, *d*z = .84] but not for the line-drawings [4.15 ± .36 μV vs. 4.33 ± .35 μV, t (21) = 1.55, p = .14]. That is, the posterior P2 amplitude was related to correct classification for color photographs.

3.2.2 Anterior P2, N2, P3 and N4 effects

Figure 5 shows the ERP data at the fronto-central electrode sites, at which the ERP waveforms were computed over the group of fronto-central electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4), which were representative of the topography of each component. Figure 6 shows the latencies or amplitudes of the anterior P2, N2, P3 and N4 under each condition. Table 2 summarizes the significant results of the three-way repeated ANOVAs that were performed over the latencies or amplitudes of the anterior P2, N2, P3 and N4.

Peak latencies of the anterior P2. For the anterior P2 peak latencies, there was a significant type of image by SOA interaction. As shown in Figure 6, consistent with the edge-based theory, there was significantly shorter anterior P2 peak latency for the line-drawing trials than for the color photograph trials for all SOAs [SOA = 13 ms: 176.39 ± 3.72 ms vs. 171.19 ± 4.44 ms, t (21) = 2.13, p < .05, *d*z = .45; SOA = 27 ms: 181.97 ± 3.40 ms vs. 171.81 ± 4.45 ms, t (21) = 3.02, p < .01, *d*z = .64; SOA = 40 ms: 193.01 ± 3.93 ms vs. 171.25 ± 4.46 ms, t (21) = 6.02, p < .001, *d*z

= 1.28]. Importantly, the P2 peak latency significantly increased with the SOA for the color photograph trials (*ps* < .05) but not for the line-drawing trials (*p*s > .84). Similarly, the linear trend was 0 ms (*SE* = 3.8 ms) for the line-drawings; using the uniform [0, 17] to represent the alternative (where 17 was the linear trend for the color photographs) yields *B* = 0.28, which is also substantial evidence for the null hypothesis. Thus, consistent with the results of posterior P1 and N1latencies, the results of anterior P2 latencies provided strong evidence for the edge-based theory.

In addition, the type of image by correctness interaction was significant. The P2 peak latency was significantly shorter in incorrect than correct trials for the color photographs [176.62 ± 2.57 ms vs. 190.95 ± 4.85 ms, t (21) = -4.16, $p < .001$, dz = .89], but not for the line-drawings [172.07 ± 3.93 ms vs. 170.76 ± 4.53 ms, t (21) = -.51, p = .62]. The anterior P2 latency difference between the color photograph trials and line-drawing trials was similar to the posterior N1 peak latency.

Peak amplitudes of the anterior P2. For the anterior P2 peak amplitudes, there was a significant type of image by correctness interaction. The anterior P2 peak amplitude was significantly larger for incorrect than correct trials only for the color photographs [6.53 ± .80 μV vs. 5.62 ± .68 μV, t (21) = 2.87, $p < .01$, dz = .61] but not for the line-drawings [4.45 ± .72 μV vs. 4.29 ± .73 μV, t (21) = .65, p = .53].Moreover, for both correct and incorrect trials, the anterior P2 amplitude was significantly larger for the color photograph trials than for the line-drawing trials (*p*s < .01). That is, the anterior P2 amplitude difference between the incorrect and correct trials was similar to the posterior P2, while the anterior P2 amplitude difference between the color photographs and line-drawings was opposite to the posterior P2.

Amplitudes of N2, P3, and N4. As shown in Figure 6, consistent with our prediction, the three-way ANOVA on N2, P3, and N4 revealed only significant main effects. The amplitudes of N2, P3, and N4 were all significantly larger for the color photograph trials than for the line-drawing trials (all *p*s < .001). However, for both types of images, a longer SOA led to significantly decreased effects of N2, P3, and N4 (all *p*s < .05), while incorrect trials of both types of images elicited significantly greater N2 but smaller N4 effects (both *p*s < .05). The results indicated that the later components varied with the SOA and correctness similarly for the two types of images.

3.2.3 The relationship between the behavioral accuracy and ERP effects

To further explore the relationship between the accuracy rates and latencies or the amplitudes of the ERP components, the accuracy rates were stepwise regressed on the incorrect-correct difference for the latencies or amplitudes of all of the components (i.e., incorrect minus correct latency or amplitude of each component averaged over the SOAs) separately for the color photographs and line-drawings. For the color photographs, this step revealed a relationship between the accuracy rates and peak latencies of the anterior P2 which reached only marginal significance, F (1, 20) = 4.04, p = .058, R^2 = .17. For the line-drawings, this step revealed two significant models: (1) the amplitude differences of the anterior N2 significantly predicted the accuracy rates, F (1, 20) = 5.78, p = .026, R^2 = .22; (2) the amplitude differences of the anterior N2 and P2 significantly predicted the accuracy rates, F (2, 19) = 7.99, p = .003, R^2 = .46. Thus, the anterior P2 latency appears to be an indicator of the accuracy for the color photographs, while the anterior N2 and P2 amplitudes appear to be indicators of the accuracy for the line-drawings.

4. Discussion

Our behavioral results showed that the correct classification was higher for the color photograph trials than for the line-drawing trials when the SOA was longer than 13 ms, but crucially, it was lower when the SOA was 13 ms. These results reconcile the apparently contradictory empirical findings of Bie~~i~~derman and Ju (1988) with those of Wurm et al. (1993) and Goffaux et al. (2005), and are consistent with our prediction that the role of surface information is modulated by the processing duration. Specifically, when the processing time was extremely limited, the color and other surface properties impaired rather than improved the performance on the color photograph trials; even when the processing time was longer, the contribution of the surface-based information to accuracy was very limited and much smaller than that of the edge-based information. The results provided new behavioral evidence for the edge-based theory which assumes that the edge-based information determines primarily performance in visual recognition and gets priority processing.

Importantly, if edge-based information receives the first analysis and the surface-based information is analyzed as the second route for recognition, then we predict that the latencies of early components that are sensitive to elemental features of stimuli would be faster for the line-drawing trials than for the color photograph trials. Previous studies revealed that the posterior P1 is the first component that indicates the spatial selective attention and the posterior N1 and the anterior P2 are associated with feature detection or integration (Hillyard & Münte, 1984; Luck & Hillyard, 1994). Thus, we analyzed the peak latencies of the posterior P1, N1, and the anterior P2 components. Consistent with the prediction, our ERP results revealed that most latencies of the posterior P1, N1, and the anterior P2 were faster for the line-drawing trials than for the color photograph trials. Specifically, the results showed that the posterior P1 peak latency was faster for the line-drawing trials than for the color photograph trials when SOA was 40 ms, the posterior N1 peak latency was faster for the line-drawing trials than for the color photograph trials when SOA was 27 and 40 ms, and the anterior P2 peak latency was faster for the line-drawing trials than for the color photograph trials when SOA was 13, 27, and 40 ms. Nonetheless, there was a slower P1 peak latency and a similar N1 peak latency for the line-drawing trials compared to the color photograph trials when SOA was 13 ms, and a similar P1 peak latency when SOA was 27 ms. Crucially, an increase in the SOA produced an linear increase in the latencies of all the three components for the color photograph trials but not for the line-drawing trials. The absolute increase value of the latency for the color photograph trials tended to rise up as one from the posterior P1 (12.00 ms) to N1 (17.56 ms) or the anterior P2 (16.62 ms). Thus, the results indicated that the shorter SOA was sufficient for extracting usable information from line-drawings, whereas more usable information continued to be extracted from color photographs as the SOA increased, which was consistent with the edge-based theory.

Moreover, incorrect trials elicited shorter latencies of the posterior N1 and the anterior P2 compared to correct trials for color photographs but not for line-drawings, indicating that incorrect categorization of color photographs may arise from insufficient processing time of extracting relevant information from color photographs. Coincidently, the regression results revealed that the accuracy rates for the color photograph trials instead of line-drawing trials could be predicted by the anterior P2 latency, suggesting that a longer anterior P2 latency is related to the higher accuracy rate for color photographs. That is, the results confirmed that sufficient processing time was crucial for extracting useful information from color photographs. This also explains why people performed worse on the color photograph trials than on the line-drawing trials when the processing time was extremely limited, i.e., when the SOA was 13 ms.

As there was such a short variable SOA (i.e., 13, 27, and 40 ms) and long-duration mask (100 ms), ~~the ERPs reflected the neural responses to an integrated target-plus-mask signal. Thus,~~ it is possible that the latencies of the early components reflected the processing of the target plus the mask with different onset time. –Nevertheless, previous neurophysiologic studies in monkeys, using line segments as stimuli, have demonstrated that backward masking typically does not have significant~~no~~ effect on the latencies of the early components in early visual areas (see Lamme, Zipser, & Spekreijse, 2002). Consistently, our results revealed that the latencies of the early components did not change with the SOA for the line-drawings. But we also found that the latencies of the early components gradually increasing with the SOA for the color photographs. As the mask onset time is identical for line-drawings and color photographs, the different latency patterns between the two conditions could not be due to the processing of the mask but the processing of the target image. That is, surface-based information involved in color photographs is not processed simultaneously with edge-based information, which is consistent with the edge-based theory.

Then, ~~However,~~ why would the edge-based information of the color photographs not be processed in the same way as that of the line-drawings, especially when SOA was 13 ms? There are at least two possible explanations: either because the edge-based information in the color photographs was not present to the same degree in the line-drawings (due to lower contrast for example), or the presence of surface information influenced the processing of edge-based information. The former explanation is consistent with the edge-base theory, while the later one is in favor of an early mechanism for surface detection, which seems inconsistent with the edge-based theory. However, it should be noted that the low performance for color photographs than for line-drawings when SOA was 13 ms indicated that this possible early detection of surface properties did not lead to early facilitation effects. That is, although there is possibly an early mechanism for surface detection, surface properties are still less efficient routes for accessing the memorial representation in natural scene categorization, which is consistent with the edge-base theory.

Previous studies have shown that the magnocellular (M) pathway (which is sensitive to the luminance contrast) is faster than the parvocellular (P) pathway (which is sensitive to the chromatic contrast and generally less sensitive to the luminance contrast) (Baseler & Sutter, 1997). The color photographs contained both luminance and chromatic information, while the line-drawings contained only luminance information; thus, our findings are consistent with the previous findings. Moreover, in Bar’s model, it is argued that low spatial frequencies (i.e., the global features of the image) conveyed by the M pathway are perceived earlier than high spatial frequencies (i.e., the fine properties of the image) (Bar, 2003; Schyns~~,~~ & Oliva, 1994). This relationship has been supported by a number of studies. For example, it is found that the inferior temporal cortex responded to low spatial frequencies 51 ms earlier than when it received high spatial frequencies (Sugase~~ et al.~~, Yamane, Ueno, & Kawano, 1999). Low spatial frequency information represents global information about the shape (Bar, 2003) or reveals salient information about the global scene structure (Schyns~~,~~ & Oliva, 1994). Although line-drawings are famous for conveying high spatial frequency information while blurry blobs are known to convey lower spatial frequency information, the global structure in the line-drawings produced by trained artists tracing the outlines was more salient than that in the color photographs. Thus, our findings are also partially consistent with Bar’s model (Bar, 2003).

Nonetheless, our findings appear to be inconsistent with the finding that color can be perceived earlier than form (Moutoussis & Zeki, 1997). In this previous study, colors were presented on the right half of the screen and oriented lines on the left half of the screen. Both the colors and lines switched with a square-wave oscillation, and the participants were asked to report what the color was when the bars tilted to the right or left. The perception in their study was conscious. In our study, the stimulus was presented for 13 ms with a variable SOA of 13, 27, 40 ms between the image and the mask. Due to the limited processing time, the perception in our study was mainly unconscious subjectively. It has been argued that form or contour processing proceeds faster than surface processing at the unconscious level such as V1 and, by contrast, surface processing proceeds faster than form or contour processing at the conscious level (Breitmeyer & Tapia, 2011). Crucially, the early peak latencies that are within 200 ms after the stimulus onset reflect unconscious processing as a precursor to conscious perception and not a separate pathway. In other words, although the contour usually receives priority processing in early scene analysis, this circumstance need not imply that the reaction time is faster for the contours than for color in conscious perception.

Surprisingly, although the anterior P2 amplitude was greater for the color photographs than for the line-drawings, the posterior P2 amplitude was larger for the line-drawings than for the color photographs. Enhanced anterior P2 has been found when people attend to a specific color (Hillyard and Münte, 1984) or when only one of several simultaneously presented objects contains the target feature (Luck & Hillyard, 1994), which indicates that the anterior P2 reflects the detection of a specific feature with feature-based attention (p. 331-332, Luck, 2012) or top-down matching processes (Evans & Federmeier, 2007). Increased posterior P2 has been found when the targets are preceded by non-informative cues rather than valid and invalid cues, which suggests that the posterior P2 reflects relatively late processing of the stimuli in the visual areas (Talsma ~~et al.~~, Slagter, Nieuwenhuis, Hage, & Kok, 2005). Because the posterior P2 amplitude gradually decreased with the SOA for both types of images and it was greater for incorrect than correct trials for the color photographs, the posterior P2 amplitude might reflect a top-down redetection or filling-in of features (Komatsu, 2006; Paradiso et al., 2006) in the early visual areas.

Unlike the above components, for both types of images, the effects of N2, P3, and N4 at the frontocentral sites gradually decreased with the SOA, despite the effects being larger for color photographs than for line-drawings. Because the N2 reflects an actively attended mismatch between a stimulus and a mental template while the P3 appears to reflect top-down monitoring by frontal attention mechanisms that are engaged in evaluating incoming stimuli (see Folstein & Van

Petten, 2008 for review; Polich, 2007), the results were consistent with decision-making becoming easier with longer SOAs. Moreover, incorrect trials of both types of images elicited greater N2 and smaller N4 effects. The later components varied with SOA and correctness similarly for color photographs and line-drawings, which is in agreement with the prediction that edge-based representation is sufficient for decision making.

Finally, we should note that there were some limitations in the present study. First, we did not include grayscale images in the study, and thus, we could not differentiate the roles of color and other surface properties such as brightness and texture in rapid natural scene categorization. Future research should explore this arrangement by comparing the grayscales with color photographs and line-drawings separately. Second, we did not manipulate the luminance contrasts and spatial frequencies of the color photographs and line-drawings in the study. Further studies should investigate this type of scenario by keeping the color photographs and line-drawings at similar luminance contrasts or spatial frequencies. Third, to compare the ERPs elicited by edge-based information and surface-based information, we used color masks for color photographs and gray masks for line-drawings. Further research should examine the role of different type of masks in ERPs for color photographs and line-drawings.

To summarize, our behavioral and ERP results provide converging evidence that edge-based information receives priority processing and plays a crucial role in natural scene categorization, whereas surface information can help to improve judgment only when the processing duration is sufficient. These results reconcile the apparently contradictory empirical findings and theoretical predictions by the edge-based and surface-based theories and help us to understand the role of edge-based and surface-based information in rapid scene categorization and how the human brain categorizes different visual stimuli in natural scene categorization.

Acknowledgments

The authors thank Prof. Dirk B. Walther in the Department of Psychology, the Ohio State University, for providing us the color photographs and line-drawings that were first used in Walther et al. (2011). The authors thank Kaiyun Li and Tinting Li for running the experiment, and we thank Prof. Kan Zhang at the Institute of Psychology, Chinese Academy of Sciences, for allowing us to use the NeuroScan equipment of his lab. This work was supported by the Natural Science Foundation of China (31270024, 61322206).

Author Contributions

QF, YJL, WC, and XF designed the experiment, QF, YJL, and WC prepared materials and performed the experiment, QF, ZD, and JW analyzed the data, and QF, ZD, JW, and XF wrote the paper.

References

Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorization. *Vision Research*, *45*, 1459–1469.

Bar, M. (2003). A cortical mechanism for triggering topdown facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*, 600–609.

Baseler H. A., & Sutter, E. E. (1997). M and P components of the VEP and their visual field distribution. *Vison Research, 37*, 675–690.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Biederman, I., & Ju, G. (1988). Surface versus-edge-based determinants of visual recognition. *Cognitive Psychology, 20*, 38–64.

Breitmeyer, B. G., & Tapia, E. (2011). Roles of contour and surface processing in microgenesis of object perception and visual consciousness. *Advances in Cognitve Psychology*, ***7***, 68–81.

Cave, C. B., Bost, P. R., & Cobb, R. E. (1996). Effects of color and pattern on implicit and explicit picture memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 639–653.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid ~~categorization~~ categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research, 40*, 2187–2200.

Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontier in Psychology, 5*, 781.

Evans, K. M., & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia, 45*, 1777–1790.

Feifei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, *12*, 893–924.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*, 152–170.

Gagnier, K. M., & Intraub, H. (2012). When less is more: Line drawings lead to greater boundary extension than do color photographs. *Visual Cognition*, *20*, 815–824.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., ~~et al~~& Rossion, B. (2005). Diagnostic colors contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*, 878–892.

Hansen, B. C., & Loschky, L. C. (2013). The contribution of amplitude and phase ~~spectra~~ spectra-defined scene statistics to the masking of rapid scene categorization. *Journal of Vision*, *13*, 1–21.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*, 185–198.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, *18*, 1631–1643.

Itier, R. J., Latinus, M., & Taylor, M. J. (2006). Face, eye and object early processing: what is the face specificity? *Neuroimage*, *29*, 667–676.

Joseph, J. E., & Proffitt, D. R. (1996). Semantic versus perceptual influences of color in object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 407–429.

Komatsu, H. (2006) The neural mechanism of perceptual filling-in. *Natur~~al~~e Reviews Neuroscience*, ***7***, 220–231.

Ková~~a~~cs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward-masking. *Proceedings of the National Academy of Sciences, U.S.A., 92*, 5587–5591.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14*, 1044–1053.

Laws, K. R., & Hunter, & M. Z. (2006). The impact of color, spatial resolution, and presentation speed on category naming. *Brain and Cognition, 62*, 89–97.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, U.S.A., 99*, 9596–9601.

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher-order image statistics in masking scene gist recognition. *Attention, Perception & Psychophysics, 72*, 427–444.

Loschky, L. C., Ringer, R., Ellis, K., & Hansen, B. C. (2015). Comparing rapid scene categorization of aerial and terrestrial views: A new perspective on scene gist. *Journal of Vision, 15*~~(6)~~, 1–29. doi: doi:10.1167/15.6.11.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., ~~et al~~& Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1431–1450.

Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In: S. J., Luck, & E. S., Kappenman, (Eds.). *The Oxford handbook of event-related potential components* (pp. 329–360). New York: Oxford University

Press.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*, 291–308.

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neurosci*ence, *1*, 144–149.

Moutoussis, K., & Zeki, S. (1997). Functional segregation and temporal hierarchy of the viusal perceptual systems. *Rroceedings the Royal of Society Lond Biological Science, 264*, 1407–1414.

Nieuwenhuis, S., Aston-Jones, ~~J~~G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin, 13*, 510–~~522~~532.

Otsuka, S., & Kawaguchi, J. (2007). Natural scene categorization with minimal attention: Evidence from negative priming. *Perception & Psychophysics, 69*, 1126–1139.

Paradiso, M. A., Blau, S., Huang, X., MacEvoy, S.P., Rossi, A.F.~~-~~, & Shalev, G~~et al~~. (2006). Lightness, fillingin, and the fundamental role of context in visual perception. *Progress in Brain Research, 155B*, 109–123.

Parron, C., & Washburn, D. (2010). Contrasting the edge- and surface-based theories of object recognition: Behavioral evidence from Macacques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 148–157.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neuropyhsiology, 118*, 2128–2148.

Rieger, J. W., Braun, C., Bü~~u~~lthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *Journal of Vision, 5*, 275–

286.

Rolls, E., Tovée, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*, 300–311.

Rossion, B., Kung, C-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 101*, 14521–14526.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Natual Neuroscience, 5*, 629–630.

Rousselet~~t~~, G. A., & Pernet, C. R. (2011). Quantifying the time course of visual object processing using ERPs: it's time to up the game. *Frontiers in Psychology,* doi: 10.3389/fpsyg.2011.00107.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science, 5*(4), 195–200.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.

Talsma, D., Slagter, H. A., Nieuwenhuis, S., Hage, J., & Kok, A. (2005). The orienting of visuospatial attention: An event-related brain potential study. *Cognitive Brain Research*, *25*, 117–129.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5*, 211–215.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process.

Journal of Cognitive Neuroscience, 15, 209–217.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience, 13*, 454–461.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line-drawings suffice our functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences, U.S.A., 108*, 9661–9666.

Wichmann, F. A., Sharpe, L. T., & Genenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 509–520.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 899–911.

Figure captions

Figure 1. Materials and experimental procedure. (A) Examples of six categories of color photographs and line-drawings. (B) Experimental setup and design. (C) Masks for color photographs. (D) Masks for line-drawings.

Figure 2. Accuracy rates. A) Accuracy rates for color photograph trials and line-drawing trials in each SOA, in which the dotted line was the chance level; B) Accuracy rates contributed by surface-based information, calculated by accuracy rate for color photographs minus accuracy rate for line-drawings in each SOA, and accuracy rates contributed by edge-based information, calculated by accuracy rate for line-drawings minus the chance level in each SOA; C) Awareness scores for color photograph trials and line-drawing trials in each SOA. The error bars depict standard errors.

Figure 3. ERP data at the occipital electrodes. (A) Grand-average ERPs of correct and incorrect trials for color photographs and line-drawings in each SOA, averaged across five occipital electrodes CB1, O1, Oz, O2, and CB2. (B) ERP differences of incorrect minus correct trials for color photographs and line-drawings in each SOA. (C) The scalp topography of the posterior P1, N1, and P2, incorrect minus correct trials separately for color photographs and line-drawings. (D) ERP differences of color photograph trials minus line-drawing trials for correct and incorrect ones in each SOA. (E) The scalp topography of the posterior P1, N1, and P2, color photograph trials minus line-drawing trials separately for correct and incorrect ones.

Figure 4. Latencies or amplitudes of the posterior P1, N1, and P2. (A) Latencies or

amplitudes of the posterior P1, N1, and P2 for the correct trials under each condition. (B) Latencies or amplitudes of the posterior P1, N1, and P2 for the incorrect trials under each condition. The error bars depict the standard errors.

Figure 5. ERP data at the fronto-central electrodes. (A) Grand-average ERPs of correct and incorrect trials for color photographs and line-drawings in each SOA, averaged across nine fronto-central electrodes F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4. (B) ERP differences of incorrect minus correct trials for color photographs and line-drawings in each SOA. (C) The scalp topography of the anterior P2, N2, P3 and N4, incorrect minus correct trials separately for color photographs and line-drawings. (D) ERP differences of color photograph trials minus line-drawing trials for correct and incorrect ones in each SOA. (E) The scalp topography of the anterior P2, N2, P3 and N4, color photograph trials minus line-drawing trials separately for correct and incorrect ones.

Figure 6. Latencies or amplitudes of the anterior P2, N2, P3, and N4. (A) Latencies or amplitudes of the anterior P2, N2, P3, and N4 for the correct trials under each condition. (B) Latencies or amplitudes of the fronto-central components for the incorrect trials under each condition. The error bars depict the standard errors.

Table 1. *Significant results of the three-way repeated ANOVAs performed over the latencies or amplitudes of the posterior P1, N1, and P2, considering the type of image, SOA, and correctness*.

	Posterior P1 latency		Posterior N1 latency		Posterior P1 amplitude		Posterior N1 amplitude		Posterior P2 amplitude	
	F	$\eta_p{}^2$	*F*	$\eta_p{}^2$	*F*	$\eta_p{}^2$	*F*	$\eta_p{}^2$	*F*	$\eta_p{}^2$
Typ			31.90***	.60	28.39***	.58	43.37***	.67	202.57***	.91
SOA	10.23**	.33	6.42**	.23	21.88***	.51	4.38*	.17	105.74***	.83
Acc					4.43*	.17	14.62**	.41	7.00*	.25
Typ * SOA	11.32***	.35	7.34**	.26	19.28***	.48			16.31***	.44
Typ * Acc			15.42**	.42	15.99**	.43	17.10***	.45	19.22***	.48
Acc * SOA			4.77*	.19	15.31***	.42	11.67***	.36		

Table 2. *Significant results of the three-way repeated ANOVAs performed over the latencies or amplitudes of the anterior P2, N2, P3, and N4, considering the type of image, SOA, and correctness.*

	Anterior P2 latency		Anterior P2 amplitude		Anterior N2 amplitude		Anterior P3 amplitude		Anterior N4 amplitude	
	F	η_p^2	*F*	η_p^2	*F*	η_p^2	*F*	η_p^2	*F*	η_p^2
Typ	27.66***	.57	26.07***	.55	11.57**	.36	40.35***	.66	11.77**	.36
SOA	8.95**	.30	3.30*	.14	5.79**	.22	6.39**	.23	14.98***	.42
Acc	8.21**	.28	5.80*	.22	24.35***	.54			7.26*	.26
Typ * SOA	10.53***	.33								
Typ * Acc	15.00**	.42	4.37*	.17						
Acc * SOA	3.35*	.14								